Received: 10 February 2021 | Revised: 9 August 2021 | Accepted: 15 August 2021
DOI: 10.1111/jcal.12608

ARTICLE

Journal of Computer Assisted Learning WILEY

Let me explain! The effects of writing and reading short justifications on students' performance, confidence and opinions in audience response systems

Pantelis M. Papadopoulos[1] | Nikolaus Obwegeser[2] | Armin Weinberger[3]

[1]Department of Instructional Technology, University of Twente, Enschede, The Netherlands

[2]Institute for Digital Enabling, Business School, Bern University of Applied Sciences, Bern, Switzerland

[3]Department of Educational Technology, Saarland University, Saarbrücken, Germany

Correspondence
Pantelis M. Papadopoulos, Department of Instructional Technology, University of Twente, Enschede, The Netherlands.
Email: p.papadopoulos@utwente.nl

Funding information
Aarhus Universitets Forskningsfond

Abstract

Background: The feedback offered to students in audience response systems may enhance conformity bias, while asking closed-type questions alone does not allow students to externalize and elaborate on their knowledge.

Objectives: The study explores how writing short justifications and accessing peer justifications as collective feedback could affect students' performance, confidence and opinions in multiple-choice audience response systems that apply the Peer Instruction model of voting/revoting.

Methods: For 8 weeks, 98 students, enrolled in an undergraduate course, attended each lecture following a flipped classroom approach. At the beginning of each lecture, students participated in a quiz with eight multiple-choice questions. Four of these questions included a justification form in which students could elaborate on their answers. The students were randomly grouped into two conditions according to the collective feedback they received: the Shared group ($n = 54$) could see both the percentage each question choice received from the class and the respective peer justifications, while the Unshared group ($n = 44$) could only see the percentage information.

Results: Analysis showed that students in both groups performed significantly better in questions with the justification form being available. Also, the two groups were comparable in terms of performance and self-reported level of confidence suggesting no main effect for making peer justification available. Despite this, students in the Shared group expressed a significantly more positive opinion in the end-of-activity questionnaire in terms of perceived learning gains and the helpfulness of writing justifications for their answers.

Take Away: Writing short justifications can have a positive impact on students' academic performance.

KEYWORDS
audience response systems, clickers, elaboration, justifications, peer instruction, self-assessment

1 | INTRODUCTION

1.1 | Benefits of ARS

Audience response systems (ARS; aka clickers) have been used widely in Education, supporting a range of instructional needs (for an overview, see Chien et al., 2016; Han, 2014; Hunsu et al., 2016; Kay & LeSage, 2009). Quiz-based activities can be designed for individual or group work (McDonough & Foote, 2015) and used at different times before, within, or after a lecture to identify preconceptions and assumptions (Anderson et al., 2013; Hoekstra & Mollborn, 2012).

Orchestrating lectures with interactive learning arrangements, such as discussions, exercises and collaborative sessions is challenging (Gehlen-Baum et al., 2014). ARS are tools for supporting interactive lectures, to activate and challenge students' prior knowledge, to engage students in critical thinking and reflection, and ultimately, to foster understanding and enhance retention (Wu et al., 2019). Typically, ARS allow the teacher to broadcast a question to the whole class simultaneously, tally student answers (i.e., take in collective feedback; see Michinov et al., 2020), and provide immediate and personalized feedback to the student (Chien et al., 2016). An ARS-based quiz can employ different forms of closed-type items such as true/false questions, fill-in-the-blanks exercises, sorting-list items, multi-selection questions and, of course, multiple-choice questions (Desrochers & Shelnutt, 2012; Sutherlin et al., 2013). The fact that feedback in ARS is usually anonymous, provides an additional level of psychological safety to the students and increases the acceptance of the process (Barr, 2017; Bojinova & Oigara, 2013; Stowell et al., 2010).

Studies on the pedagogical benefits of using ARS in the classroom have reported higher student engagement (Crouch & Mazur, 2001; Poirier & Feldman, 2007), increased student satisfaction (Marshall et al., 2012) and development of critical thinking (Mollborn & Hoekstra, 2010). Regarding academic performance, empirical evidence has linked ARS activities with better knowledge acquisition (Mayer et al., 2009; Shapiro et al., 2017; Shapiro & Gordon, 2012, 2013), higher course grades (Brady et al., 2013; Mayer et al., 2009; Poirier & Feldman, 2007) and improved retention (Prince, 2004). Finally, as far as teachers' acceptance of ARS is concerned, using quizzes in lectures poses a low technological barrier (Blackwell et al., 2013; Ertmer et al., 2012) with studies reporting on teachers' positive attitudes towards ARS in identifying misconceptions and assisting in organizing lectures (Chen et al., 2010).

1.2 | Limitations of ARS

Despite the recognized educational value of ARS, there are also studies that showed inconclusive (Elicker & McConnell, 2011) or even detrimental outcomes from using ARS (Fortner-Wood et al., 2013), thus suggesting that the tool itself is not a panacea for changing the teaching and learning practices in lecture halls. Even though Shapiro (2009) argued that factors related to methodology and the technology used may have played a role in the negative outcomes recorded in some studies, many researchers suggest a need for further studies on ARS (Chien et al., 2016; Han, 2014; Kay & LeSage, 2009) and their underlying pedagogy (Kennedy & Cutts, 2005; Mun et al., 2009; Shapiro et al., 2017). Consequently, a first point of consideration is, how ARS extend the learning affordances within a lecture and how that aligns with appropriate pedagogies.

Focusing on the affordances of the technology, we can identify two major limitations of current ARS. First, the feedback that is offered to students in most ARS is the distribution of anonymous answers to each question choice (either as percentage values or size of student subgroups). This information alone does not provide insights into the characteristics of the audience such as their prior knowledge and reasoning. Consequently, students may feel encouraged to focus more on gaming the system strategies, thus exhibiting conformity bias and changing their initial answers to the most popular one (Baker et al., 2013; Michinov et al., 2020; Nielsen et al., 2012; Perez et al., 2010).

Second, asking multiple-choice questions alone via ARS is no warrant for facilitating student externalization and elaboration of knowledge, but a question of how ARS use is embedded and orchestrated with other phases for revisiting the topic and the collective feedback received.

Mazur's, widely adopted, voting/revoting Peer Instruction model (Crouch & Mazur, 2001; Mazur, 1997, 2009) for ARS activities defines a sequence of steps of how to engage students in dealing with the collective feedback:

1. Provide their initial answers to multiple-choice questions (aka voting phase).
2. Receive collective feedback based on class responses through the ARS.
3. Discuss their answers with a neighbouring peer.
4. Answer the same questions for a second time (aka revoting phase) before they receive the correct answers and participate in the teacher-led class discussion that follows.

Therefore, students can elaborate on their answers during the peer discussion phase. Peer Instruction has been successful in improving student/student and student/teacher interaction in lectures (Blasco-Arcas et al., 2013) and has been associated with higher student performance (Crouch et al., 2007; Crouch & Mazur, 2001; Mayer et al., 2009). However, allocating the necessary time for peer discussion is not always efficient within the restricted timeframe of a lecture, especially in cases of large audiences and quizzes with multiple questions. As a result, studies on ARS have employed a range of variants and extensions of the Peer Instruction model (for an overview, Balta et al., 2017; Vickrey et al., 2015).

1.3 | Confidence and knowledge elaboration in audience response systems

To address the conformity bias in ARS activities, researchers have suggested enriching the collective feedback students receive after the voting phase with metrics that would provide more information on the audience. For example, studies have investigated how eliciting metacognitive judgements of confidence on the correctness of an answer and providing collective confidence information could support reflection and self-regulation (Kleitman & Costa, 2014; Schnaubert & Bodemer, 2015, 2017, 2019). Asking students to assess their levels of confidence, preparation, understanding, and so on, has been used extensively in learning settings as a way to analyse students' metacognition and self-assessment accuracy (Schraw, 2009; Veenman, 2017). As Soderstrom et al. (2015) suggested, even the act of eliciting such assessments has the potential of making the study material more memorable and increasing performance. Metacognitive judgements of confidence, in particular, have been deemed strong predictors of academic achievements (Stankov et al., 2013). Following this rationale, we investigated in a previous study (Papadopoulos et al., 2019) two metacognitive judgements (feeling of preparedness for the upcoming quiz, and confidence in the correctness of each answer) and two objective metrics retrieved from the ARS tool (percentage of students that voted each question choice, and their past performance in course quizzes). The findings of that study showed that the enriched collective feedback and especially the confidence judgements resulted in better student performance while diminishing the conformity bias. Nevertheless, students are often inaccurate in self-assessing (Veenman, 2017), and as Michinov et al. (2020) suggested, students should be guided and made aware that collective feedback may not be accurate.

Eliciting and sharing written justification can provide an additional dimension to the collective feedback. We hypothesize that this will allow students to better understand the peers' perspectives and be more critical towards collective feedback. In addition, articulating their reasoning can help students elaborate and reflect their understanding (Papadopoulos et al., 2009, 2011; Lachner et al., 2021; Nückles et al., 2009, 2020). According to Menary (2007, p. 622), 'creating and manipulating written sentences are not merely outputs from neural processes but, just as crucially, they shape the cycle of processing that constitutes a mental act'. One issue with peer discussion in PI is that it removes anonymity since students must openly discuss their answers with their peers. Consequently, students often appear hesitant in participating in peer discussions during a quiz activity (Michinov et al., 2015). Enhancing the voting with written justifications will allow students to elaborate on their reasoning anonymously, which in turn can guide peers in assessing the choices made based on the reasons given.

2 | STUDY MOTIVATION AND RESEARCH HYPOTHESES

In our series of studies on ARS, we focus on the Peer Instruction paradigm with a notable difference: the substitution of the peer discussion between the voting and revoting phase with enriched collective feedback information that, while remaining to be anonymous, could provide students with a clear enough picture to understand their peers' profile and positions.

Based on the theoretical considerations discussed above, the study tested the following null hypotheses on whether writing justifications in the voting phase and having access to peer justifications in the revoting phase affects students' performance and confidence:

H_{01}(j-p): Writing short texts to justify answers in multiple-choice questions does not affect students' performance in the voting phase.

H_{02}(j-c): Writing short texts to justify answers in multiple-choice questions does not affect students' perceived level of confidence in the voting phase.

H_{03}(f-p): Making peer justifications available as collective feedback does not affect students' performance in the revoting phase.

H_{04}(f-c): Making peer justifications available as collective feedback does not affect students' perceived level of confidence in the voting phase.

Since the students were not familiar with the ARS tool we employed in the activity and had no experience in the PI model, the study also explored the following research question:

RQ: How do students of the two study conditions evaluate the activity in terms of (a) technology usefulness and usability, (b) learning gains, (c) helpfulness of writing justifications and (d) receiving collective feedback?

3 | METHOD

3.1 | Participants

The study was conducted as an optional activity within the 'Business Development with Information Systems' course, a second-year course with a cohort of 161 students. The students were grouped randomly by the audience response system into two conditions. However, the analysis was based only on the 98 students (age: $M = 21.15$, $SD = 3.10$) that participated in all parts of the study.

- Shared: 54 participants, who were to provide justifications that were shared with their peers.
- Unshared: 44 participants, who were to provide justifications, but whose justifications were not shared with their peers.

3.2 | Material and instruments

3.2.1 | The SAGA tool

SAGA (acronym for self-assessment/group awareness), the ARS tool used in the study, was developed by our team, is based on the vote/revote model, allows the teacher to define the feedback metrics used in the activity, and has been used in several courses and studies.

Answering a question during the voting phase of the current study included three actions: (a) answering the multiple-choice question, (b) writing a short justification (up to 140 characters) explaining why a choice was selected and (c) denoting the level of confidence that the answer was correct (Figure 1). The justification form could

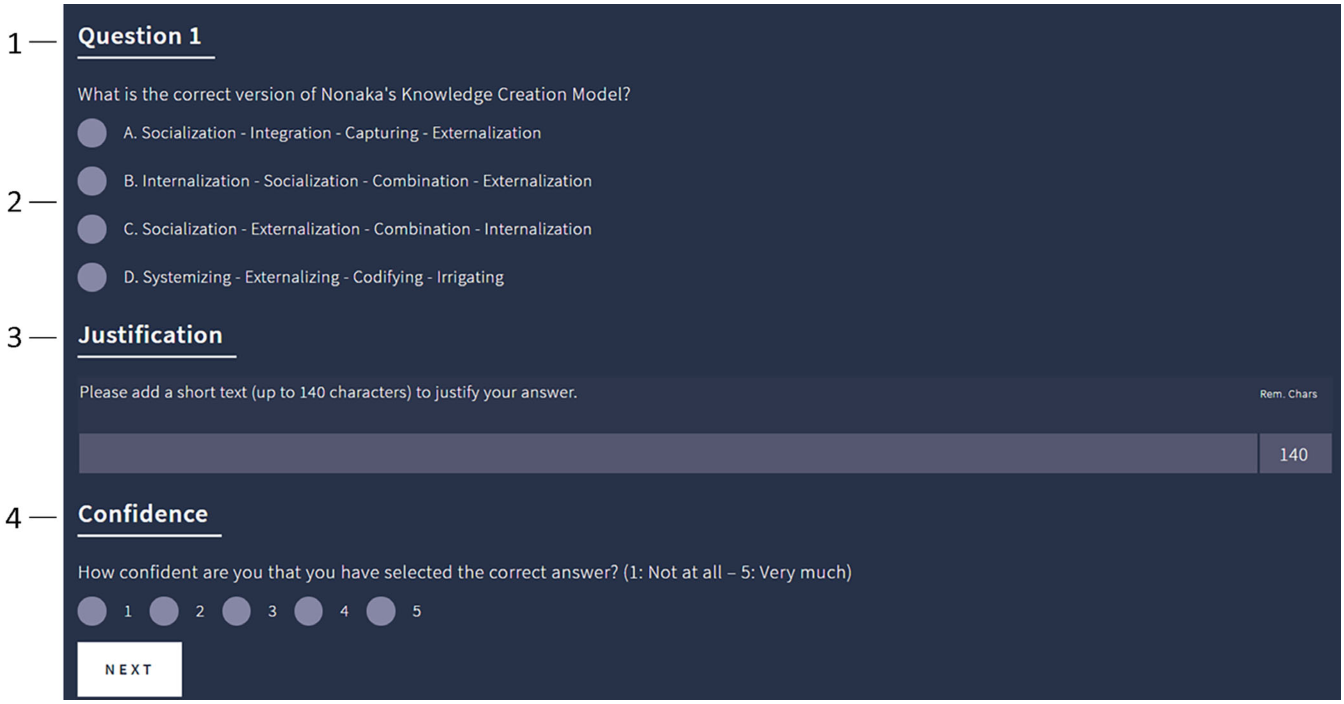

FIGURE 1 SAGA screenshot during the voting phase for questions with the justification field. 1: Question description; 2: Four choices were offered for each question; 3: Justification form—Up to 140 characters allowed; 4: Confidence form using a five-step Likert scale—The confidence data were transformed into percentage values in the analysis [Colour figure can be viewed at wileyonlinelibrary.com]

not be submitted empty and students had to type at least one character in the form to proceed. We considered all nonsensical answers (e.g., ' ', '.', 'asdf') as empty in our analysis. Also, the students were not able to go back to an answered question nor to skip ahead to the next question without answering the current one.

Adding a writing task in a tool that was previously used for clicking only poses a few design challenges as formulating and writing elaborations requires additional time. Regarding time constraints, the whole quiz activity had to be short not to disrupt the lecture plan (a common critique for quiz activities—Koenig, 2010; Strasser, 2010). Limiting the length of justifications (140 characters were allowed) helped to address this issue and emphasized to students that indeed only short justifications were needed. The short justifications served lowering the threshold for students to go through a varied set of justifications, thus getting to know a broad scope of peer perspectives within the timeframe of the activity.

In the revoting phase, the students were able to change their initial answer and confidence level, while collective feedback (i.e., percentage information and peer justifications) was presented to them according to their study condition (Figure 2).

3.2.2 | The end-of-activity questionnaire

After the last quiz activity in the study, we asked students to fill in an evaluation questionnaire addressing different aspects of their experience. The students had to denote their agreement to a set of statements using a five-step Likert scale (1: strongly disagree–5: strongly agree). The first part of the questionnaire focused on evaluating the usability and usefulness of SAGA (e.g., ease-of-use, enjoyment, engagement), helping our team identify design issues we would need to address. The second part focused on the perceived learning gains (e.g., understanding concepts, identifying misconceptions, self-assessing performance). The third part focused on the impact writing justifications had on students during the voting phase, and the last part focused on the helpfulness of the received collective feedback in the revoting phase (i.e., percentage for the Unshared group; percentage and peer justifications for the Shared group). The questionnaire was administered within SAGA and students could use their system credentials and answer it anonymously.

3.3 | Procedure

The course lasted 14 weeks and the study started on the second and finished on the ninth week for a total of 8 weeks. During the first week, the lecture focused on introducing the course, the SAGA tool, and the planned activities. The students were informed of the research nature of the activity and the fact that some of them will not see the justifications that others wrote during the revoting phase. The first quiz was administered in the lecture of the second week, while after the quiz on the ninth week, the students filled in the end-of-activity questionnaire. The activity was not linked to the formal course assessment and participation was completely anonymous (student activity was monitored using a system-generated id). As an extra incentive for meaningful participation, we informed the students at the beginning of the study that gift vouchers for the university's bookstore (valued at 10 euros/12 USD) will be given to the five students with the highest overall scores of each condition (these 10 students had to reveal their identities to receive the vouchers).

Figure 3 presents an overview of the study and the procedure within each quiz activity. The course design implemented aspects of the flipped classroom paradigm, as each lecture's material (e.g., slides, articles, online resources) was available a week in advance, and students were encouraged to study it before coming to class. During the first 15 min of each lecture, the students participated in the ARS quiz (10 min for the voting and 5 min for the revoting phase). After that, the correct answers were shown and discussed by the course instructor, while the planned lecture was following.

A weekly quiz included eight questions with only four of them (appearing in random positions within the quiz) having the form that allowed students to write justifications. In the other four, the justification form was not available and students had to only answer the question and state their confidence. During the revoting phase, students in the Unshared group were able to see the percentage each question choice received from the class, while students in the Shared group

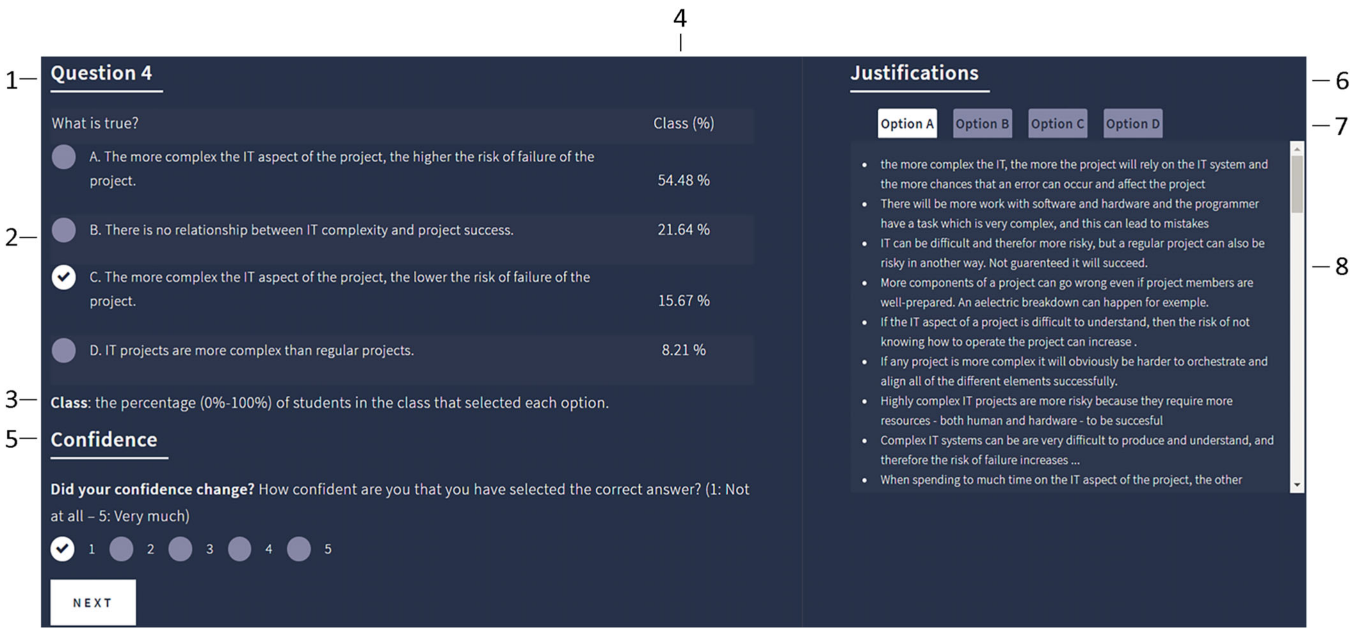

FIGURE 2 SAGA screenshot during the revoting phase for the shared group. The justification tab was hidden for the unshared group. 1: Question description; 2: The four choices and student's answer in the voting phase—The student could change or submit the same answer; 3: Description of the percentage feedback metric; 4: The percentage feedback metric; 5: The confidence form and student's confidence level in the voting phase—The student could change or submit the same confidence level; 6: Peer justifications panel—Visible only for the shared group; 7: Peer justifications tabs—By clicking on each tab, the student was able to see the peer justifications for the respective question choices; 8: Peer justifications sorted by length [Colour figure can be viewed at wileyonlinelibrary.com]

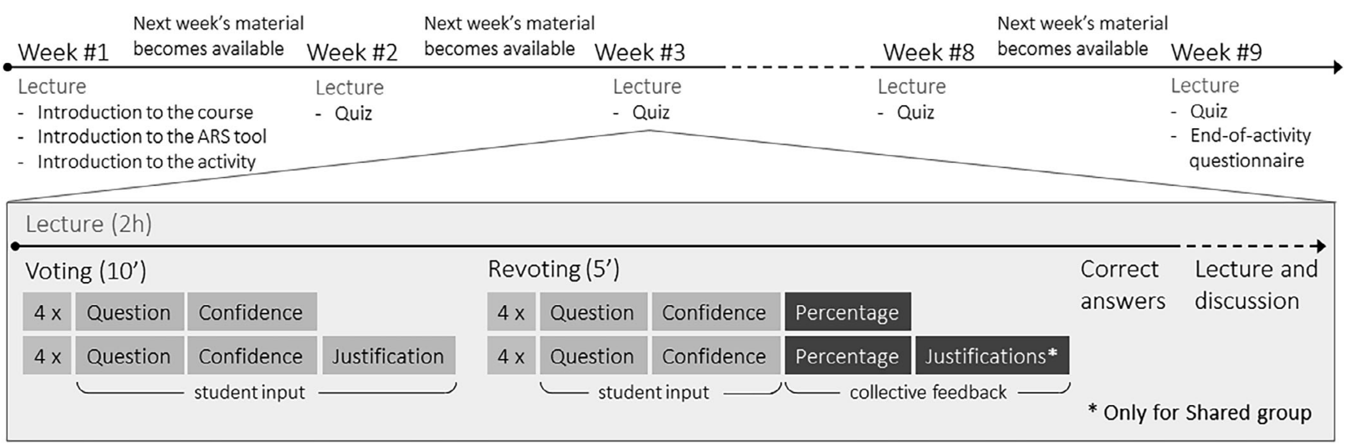

FIGURE 3 Study overview

were able to see, in addition to the percentage, all the submitted justifications for each one of the question choices, sorted based on their length (in characters) so that more elaborate justification would appear on top of the list without the need for scrolling.

3.4 | Measures

The study included two independent variables based on the study conditions in the voting (availability of the justification form) and revoting phase (availability of peer justifications). Students' performance, confidence and opinions as recorded in the end-of-activity questionnaire were the dependent variables of the study. Table 1 presents each dependent variable along with the actual scale and the way the scale was transformed and used in the analysis. Performance and confidence were calculated as percentages in the voting phase, while Hake's (1998) normalized gain (<*g*>) was used to assess the difference between voting and revoting (1).

$$\langle g \rangle = \frac{\text{revoting}(\%) - \text{voting}(\%)}{100 - \text{voting}(\%)} \tag{1}$$

Table 2 shows how the dependent variables were operationalized to address the hypotheses and research question of the study. For all statistical analysis $\alpha = 0.05$.

4 | RESULTS

4.1 | Student performance and confidence

All students elaborated on their answers by providing justifications for at least some of the questions (Length in characters: Shared: $M = 50.18$, $SD = 25.44$; Unshared: $M = 48.48$, $SD = 28.03$), while the total usage of the justification form was 76% (2383 justifications submitted out of the possible 3136—i.e., 98 students × 32 questions with justifications). Both groups scored higher in the question subset that included the justification form (Perf.Init.NoJ: Shared: $M = 50.81$, $SD = 21.56$; Unshared: $M = 49.50$, $SD = 21.38$; Perf.Init.Just: Shared: $M = 64.88$, $SD = 19.19$; Unshared: $M = 60.50$, $SD = 18.69$), while paired-samples *t*-tests showed that this difference was significant for both groups (Shared: $t[53] = 3.35$, $p < 0.01$, $d = 0.65$; Unshared: $t[43] = 3.46$, $p < 0.01$, $d = 0.75$).

Similarly, both groups felt more confident about their answers in questions where they were able to write justifications (Conf.Init.NoJ: Shared: $M = 54.23$, $SD = 16.01$; Unshared: $M = 56.42$, $SD = 16.63$; Conf.Init.Just: Shared: $M = 58.68$, $SD = 15.67$; Unshared: $M = 59.81$,

TABLE 2 Dependent variables operationalization

Hypothesis/RQ	Operationalization
H_{01}(j-p)	Paired-samples *t*-test for Perf.Init.NoJ and Perf.Init.Just for the two groups
H_{02}(j-c)	Paired-samples *t*-test for Conf.Init.NoJ and Conf.Init.Just for the two groups
H_{03}(f-p), H_{04}(f-c)	One-way multivariate analysis of variance (one-way MANOVA) using the availability of peer justifications in the collective feedback as the independent variable (Shared/Unshared) and Perf.Gain.Just and Conf.Gain.Just as the dependent variables
RQ	Mann–Whitney *U* tests comparing the opinions in the two groups in Usability, Learning gains, Justifications, Feedback Wilcoxon signed-rank test comparing the opinions of the Shared group regarding the two types of collective feedback they received (i.e., percentage and peer justifications)

TABLE 1 Dependent variables of the study

Variable (scale)	Description	Analysis
Performance		
Perf.Init/Rev.NoJ (0–32)	Number of questions without the justification form ($n = 32$; i.e., four questions in each of the eight weekly quizzes) that the student answered correctly in the voting/revoting phase	%
Perf.Init/Rev.Just (0–32)	Number of questions with the justification form ($n = 32$) that the student answered correctly in the voting/revoting phase	%
Perf.Gain.Just (0–32)	Performance difference between the voting and revoting phases in the questions with the justification form	<g>
Confidence		
Conf.Init/Rev.NoJ (1–5)	Mean confidence during the voting/revoting phase in the questions without the justification form	%
Conf.Init/Rev.Just (1–5)	Mean confidence during the voting/revoting phase in the questions with the justification form	%
Conf.Gain.Just(0–5)	Mean confidence difference between the voting–revoting phases in the questions with the justification form	<g>
Opinion		
Usability (1–5)	Mean score in the end-of-activity questions related to the usability and usefulness of the ARS tool	%
Learning gains (1–5)	Mean score in the questions related to perceived learning gains	%
Justifications (1–5)	Score regarding the helpfulness of writing justifications in the voting phase	%
Feedback (1–5)	Score regarding the helpfulness of collective feedback received	%

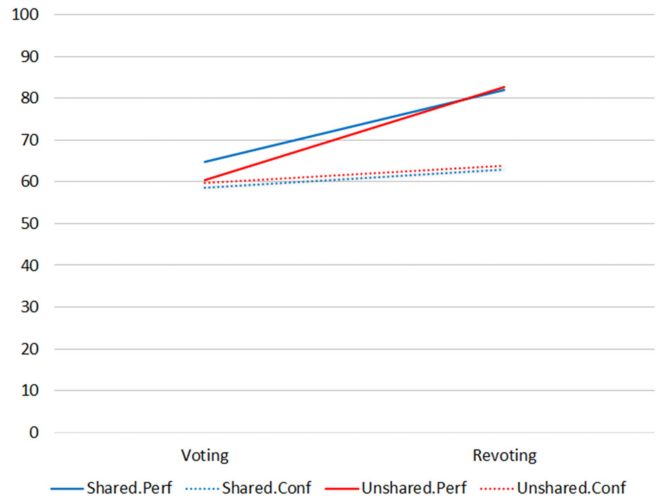

FIGURE 4 Performance and confidence for the questions with justifications [Colour figure can be viewed at wileyonlinelibrary.com]

$SD = 18.02$) and paired-samples t-test showed, once again, that this difference was significant for both groups (Shared: $t[53] = 2.88$, $p < 0.01$, $d = 0.56$; Unshared: $t[43] = 2.27$, $p = 0.02$, $d = 0.49$).

One-way MANOVA did not reveal a statistical difference between the two groups in the gains recorded (performance/confidence) in the revoting phase, $F(2, 95) = 1.47$, $p = 0.35$; Wilk's $\Lambda = 0.97$, partial $\eta^2 = 0.02$. The two groups scored comparably in the revoting phase (Perf.Rev.NoJ: Shared: $M = 65.44$, $SD = 23.31$, $\langle g \rangle = 0.30$; Unshared: $M = 65.03$, $SD = 21.09$, $\langle g \rangle = 0.31$; Perf.Rev.Just: Shared: $M = 82.03$, $SD = 23.25$, $\langle g \rangle = 0.49$; Unshared: $M = 82.63$, $SD = 20.88$, $\langle g \rangle = 0.56$) and denoted similar levels of confidence (Conf.Rev.NoJ: Shared: $M = 60.84$, $SD = 15.62$, $\langle g \rangle = 0.14$; Unshared: $M = 61.22$, $SD = 17.04$, $\langle g \rangle = 0.11$; Conf.Rev.Just: Shared: $M = 63.09$, $SD = 15.48$, $\langle g \rangle = 0.11$; Unshared: $M = 63.86$, $SD = 18.41$, $\langle g \rangle = 0.10$). Figure 4 shows the performance and confidence percentage of the two groups for the question subset with the justification form.

4.2 | Student evaluation of the activity

Table 3 shows students' responses in the end-of-activity questionnaire. Mann–Whitney U test showed that both groups expressed a similarly positive opinion about the usability and usefulness of the ARS tool (Shared: $M = 79.05$, $SD = 18.25$; Unshared: $M = 76.60$, $SD = 20.65$; $U = 574.00$, $p = 0.94$). Even though both groups also expressed positive opinions about the learning gains they experienced during the activity (Shared: $M = 76.07$, $SD = 18.73$; Unshared: $M = 67.53$, $SD = 17.87$), students in the Shared group had a significantly more positive opinion ($U = 399.00$, $p = 0.03$, $d = 0.52$). Regarding the helpfulness of writing justifications, the two groups were rather positive (Shared: $M = 76.00$, $SD = 23.00$; Unshared: $M = 65.00$, $SD = 20.00$), with the Shared group expressing, yet again, a significantly more positive opinion ($U = 404.00$, $p = 0.04$, $d = 0.51$).

TABLE 3 End-of-activity questionnaire

	Shared ($n = 54$) M (SD)	Unshared ($n = 44$) M (SD)
Usability	*79.05 (18.25)*	*76.60 (20.65)*
SAGA was easy to use	87.40 (17.00)	89.20 (20.60)
SAGA helped me get instant feedback on what I knew and did not know	83.00 (18.20)	75.80 (25.80)
Using SAGA has increased my enjoyment of lectures	71.20 (19.40)	68.60 (15.80)
Answering anonymously in SAGA encouraged me to be an active participant in the class	74.60 (18.40)	72.80 (20.40)
*Learning gains**	*76.07 (18.73)*	*67.53 (17.87)*
Using SAGA helped me to identify errors and misconceptions	73.60 (22.40)	67.20 (17.40)
SAGA allowed me to understand better the key concepts	77.00 (17.00)	67.60 (18.80)
Because of SAGA, I was more certain about how I was performing in the class	77.60 (16.80)	67.80 (17.40)
*Justifications**	*76.00 (23.00)*	*65.00 (20.00)*
During the voting phase, writing justifications helped me clarify my own understanding	76.00 (23.00)	65.00 (20.00)
Feedback	*71.50 (24.50)*	*77.20 (24.00)*
During the revoting phase, the percentage values helped me choose my final response	75.60 (21.60)	77.20 (24.00)
During the revoting phase, my peers' justifications helped me choose my final response (Shared group only)	67.40 (27.40)	

*$p < 0.05$.

Finally, all students also appreciated the helpfulness of the percentage metric in choosing their final answer in the revoting phase (Shared: $M = 75.60$, $SD = 21.60$; Unshared: $M = 77.20$, $SD = 24.00$; $U = 542.00$, $p = 0.68$). In an extra question that appeared only to the Shared group, students expressed a rather positive opinion about the helpfulness of reading peer justifications in the revoting phase ($M = 67.40$, $SD = 27.40$), but, as the Wilcoxon signed-rank test showed, this opinion was significantly less positive compared to the one about the usefulness of the percentage metric ($Z = -2.37$, $p = 0.01$, $d = 0.33$).

5 | DISCUSSION

System log files showed that all students used the justification form at a high rate. The analysis of student performance in the voting phase showed that both groups performed significantly better when they had the option of writing justifications while answering multiple-choice questions. The current empirical evidence suggests that even short elaboration exercises within ARS, in which students oriented towards average performance, can improve student learning. Therefore H_{01}(j-p) is rejected and an alternative hypothesis H_{a1}(j-p) can be stated: 'writing short texts to justify answers in multiple-choice questions has a positive effect on students' performance in the voting phase'. The findings of this study are in line with studies on how elaborating questions can enhance and support student understanding (Papadopoulos et al., 2009, 2011; Lachner et al., 2021; Nückles et al., 2009, 2020). The findings of the current study also corroborate a similar outcome on soliciting elaborated written justifications in multiple-choice questions, which showed a significant positive influence on students' choice selection (Koretsky et al., 2016).

Noting the short length of the justifications, we argue that apart from the impact elaboration may have on student performance, an alternative explanation for students' higher scores in questions with justifications could be that the modification of the answering process from 'ticking boxes' to 'ticking and typing' may have suspended gaming-the-system strategies. As the literature on gaming-the-system strategies suggests (Baker et al., 2013), one way to dissuade students from finding shortcuts and loopholes in a procedure is to make the procedure harder to game. The requirement of writing a justification may have disrupted gaming-the-system strategies, making students reflect on the question instead of mechanically clicking the 'Next' button and moving on to the next question. Of course, this needs to be investigated in future research alongside analysis of gaming detectors and answering patterns.

Analysis of students' activity in the voting phase also showed that students were more confident about their answers when they were able to write short justifications about them. As mentioned earlier, knowledge elaboration may lead to deeper understanding (Papadopoulos et al., 2009, 2011; Lachner et al., 2021; Nückles et al., 2009, 2020) and as some studies have already reported, there is a positive correlation between confidence and task performance (Atherton et al., 2014; Stankov, 2013). Based on the findings in the voting phase, H_{02}(j-c) is rejected and an alternative hypothesis H_{a2}(j-c) is proposed: 'writing short texts to justify answers in multiple-choice questions can positively affect students' perceived level of confidence in the voting phase'.

The impact of writing justifications in the voting phase on students' performance and confidence is important as the design implication to include such exercise in multiple-choice quizzes can be applied also to settings that do not follow the PI model. Additional research is needed to investigate how the writing process can be further improved and how students should be guided and encouraged to engage students in meaningful knowledge elaboration.

While writing justifications had a clear positive effect on students' performance and confidence, making peer justifications available for the Shared group was not enough for the students of that group to outperform the Unshared group. On the contrary, the Unshared group recorded a slightly higher gain during the revoting phase. Similarly, the two groups denoted comparable confidence levels, while both felt more confident in the revoting phase, thus replicating a common finding in PI studies (Papadopoulos et al., 2019; Crouch et al., 2007; Crouch & Mazur, 2001; Mayer et al., 2009). As such, hypotheses H_{03}(f-p) and H_{04}(f-c) are accepted.

A possible explanation for the absence of a main effect is that the information provided already by the percentage was strongly indicative of the correct answer and as such students in the Unshared group were able to identify the correct answer in the revoting phase. Indeed, additional feedback metrics in ARS may play a role in highly challenging questions mainly, while the percentage information does point to the correct answer in the majority of cases (Papadopoulos et al., 2019).

An alternative explanation for the comparable performance of the two groups could be that Shared students accepted the peer justification as a valid source of information without engaging in further reflection of the task. As Hattie (2012) suggested, providing feedback is necessary, but it is more important to consider how feedback is received by the students. As the analysis of the end-of-activity questionnaire revealed, the percentage values were considered highly helpful in the revoting phase. On the contrary, peer justifications were considered significantly less helpful as collective feedback, even though the Shared group expressed an overall positive opinion about them. While getting to know what others think on a subject has shown to be beneficial in peer review scenarios of learning (Papadopoulos et al., 2017; Lundstrom & Baker, 2009), simply making peer justifications accessible to students does not seem to be enough for students to actually engage and reflect on peers' justifications. Further studies are needed to explore how the visibility of peer justifications could shift from a nice-to-have feature in the student's eye to a bit of reasoning that challenges students to scrutinize their peers' and their own views.

Regarding the effect of writing justifications on how students perceived the activity, although the voting phase was identical for the two groups, students in the Shared group expressed a more positive opinion than the Unshared group regarding the helpfulness of writing justifications. The Shared group also expressed a significantly more positive opinion on the learning gains in the activity as a whole. Since both groups wrote justifications in the same way and of the same length, we argue that even though reading peer justifications as collective feedback did not improve academic performance or confidence, there is strong empirical evidence to suggest that making peer knowledge elaborations available to students can affect positively the way they perceive the activity. Writing justifications for with the peer audience in mind, leverages the authors' expectations of how their writing is being assessed by their peers, fostering their efforts (Wang et al., 2021). The motivational benefits for cognitive engagement set aside, students also simply seem to regard the activity more meaningful when the justifications have peers as addressees. This is in line with students' positive assessment of the learning experience with the ARS tool. With the impact of educational technology in authentic settings depending on the acceptance of students, it is encouraging to observe that learners reported very positively on usability and helpfulness of the ARS tool.

5.1 | Limitations

The study was conducted in vivo, that is, in an actual lecture hall and course. While this contributes to external validity of the study and feasibility of the ARS tool due to the authenticity of the setting, there are drawbacks to the study design. It is hard to distinguish, for example, between effects of writing own versus reading others' justifications. This would be the case for some students in the Shared group that did not write meaningful justifications, but they could still see peer elaborations on question choices. Similarly, it is not feasible to distinguish between effects of reading high versus low quality peer justifications. Additional experimental groups would be necessary for this line of research that would be beyond the scope of this study.

The difficulty of the questions used can be a factor in any quiz activity. To avoid the possibility of a ceiling/floor effect, we selected in the study questions from a pool that has been used and refined repeatedly over the previous years. To further check that a ceiling/floor effect did not occur, we examined the min, max, kurtosis, skewness values for the two question subsets for the two groups (Perf.Init.NoJ: Shared: $min = 27.12$, $max = 88.56$, $kurtosis = 0.41$, $skewness = 0.08$; Unshared: $min = 22.51$, $max = 83.33$, $kurtosis = 0.47$, $skewness = 0.08$; Perf.Init.Just: Shared: $min = 23.44$, $max = 81.07$, $kurtosis = 0.32$, $skewness = 0.11$; Unshared: $min = 28.90$, $max = 87.55$, $kurtosis = 0.39$, $skewness = 0.08$). Of course, we also confirmed that all question subsets yield scores under a normal distribution.

6 | CONCLUSIONS

The study provided empirical evidence on the impact that short written justifications could have on students' learning and confidence in ARS activities. Knowledge elaboration, even in the case of short justifications, seems to significantly improve academic performance. This is in line with the effects of elaboration on student learning in general (Nückles et al., 2009, 2020). At the same time, there is evidence that eliciting metacognitive judgements is based on students' level of confidence, and in several cases, such confidence is misplaced. The study also suggests that apart from elaboration, the requirement for non-empty justifications may have played an additional beneficial role in improving students' performance by disrupting gaming answering strategies.

Making peer justifications available to students as collective feedback was expected to offer an advantage to students. However, the analysis suggested that this additional information on peers' perspectives did not result in higher academic performance raising also a question about students' actual engagement with the peer justifications. Nevertheless, such information has still been deemed useful for the students as it improves their overall acceptance of the activity.

Regarding the design implications of the study, we argue that short justifications within ARS activities could provide multiple benefits for the students without posing significant overhead or delays. Adding such a functionality offers the opportunity for elaboration to all students while retaining anonymity—a crucial characteristic of ARS (Barr, 2017; Bojinova & Oigara, 2013; Stowell et al., 2010). The increased academic performance in questions with justifications is a strong advocate for adding such an affordance in ARS, while further research is needed to investigate how students could be brought to deliberately engage with peer justifications leading to higher learning gains and how technology or study design could accurately measure such engagement. With the study showing that ARS with justifications has a beneficial impact on students' performance as well as acceptance, further investigations could aim at how justifications could provide points of reference to engage students in class discussion in general. In conclusion, ARS could contribute to orchestrate a specific set of learning and teaching activities within lectures to create synergy between presenting information, engaging learners to elaborate their knowledge with ARS with justifications, and building on students' elaborations to repeat or advance a lecture.

ACKNOWLEDGEMENTS

This work has been partially funded by a Starting Grant from AUFF (Aarhus Universitets Forskningsfond), titled 'Innovative and Emerging Technologies in Education'. The authors would also like to thank Alin Panainte for his work in developing and working with the SAGA environment.

CONFLICT OF INTEREST

No author associated with this paper has any potential or pertinent conflicts which may be perceived to have an impending conflict with this work.

ETHICS STATEMENT

The study included the participation of students from Aarhus University, Denmark. The students volunteered to participate and provided their consent online in SAGA. The study was performed according to the ethics and research standards procedure that was in place at the time at Aarhus University.

DATA AVAILABILITY STATEMENT

Author elects to not share data - Research data are not shared.

ORCID

Pantelis M. Papadopoulos https://orcid.org/0000-0002-1527-5483

REFERENCES

Anderson, L. S., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (2013). The clicker technique: Cultivating efficient teaching and successful learning. *Applied Cognitive Psychology*, *27*(2), 222–234.

Atherton, S., Antley, A., Evans, N., Cernis, E., Lister, R., Dunn, G., Slater, M., & Freeman, D. (2014). Self-confidence and paranoia: An experimental study using an immersive virtual reality social situation. *Behavioural and Cognitive Psychotherapy*, *44*(1), 56–64.

Baker, R. S. J. D., Corbett, A. T., Roll, I., Koedinger, K. R., Aleven, V., Cocea, M., Hershkovitz, A., de Carvalho, A. M. J. B., Mitrovic, A., & Mathews, M. (2013). Modeling and studying gaming the system with educational data mining. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 97–116). Springer.

Balta, N., Michinov, N., Balyimez, S., & Ayaz, M. (2017). A meta-analysis of the effect of peer instruction on learning gain: Identification of informational and cultural moderators. *International Journal of Educational Research*, *86*, 66–77.

Barr, M. L. (2017). Encouraging college student active engagement in learning: Student response methods and anonymity. *Journal of Computer Assisted Learning*, *33*(6), 621–632.

Blackwell, C. K., Lauricella, A. R., Wartella, E., Robb, M., & Schomburg, R. (2013). Adoption and use of technology in early education: The interplay of extrinsic barriers and teacher attitudes. *Computers & Education*, *69*, 310–319.

Blasco-Arcas, L., Buil, I., Hernández-Ortega, B., & Sese, F. S. (2013). Using clickers in class. The role of interactivity, active collaborative learning and engagement in learning performance. *Computers & Education*, *62*, 102–110.

Bojinova, E., & Oigara, J. (2013). Teaching and learning with clickers in higher education. *International Journal of Teaching and Learning in Higher Education*, *25*(2), 154–165.

Brady, M., Seli, H., & Rosenthal, J. (2013). “Clickers” and metacognition: A quasi-experimental comparative study about metacognitive self-regulation and use of electronic feedback devices. *Computers & Education*, *65*, 56–63.

Chen, J. C., Whittinghill, D. C., & Kadlowec, J. A. (2010). Classes that click: Fast, rich feedback to enhance students' learning and satisfaction. *Journal of Engineering Education*, *99*(2), 158–169.

Chien, Y.-T., Chang, Y.-H., & Chang, C.-Y. (2016). Do we click in the right way? A meta-analytic review of clicker-integrated instruction. *Educational Research Review*, *17*, 1–18.

Crouch, C. H., Watkins, J., Fagen, A. P., & Mazur, E. (2007). Peer instruction: Engaging students one-on-one, all at once. *Reviews in Physics Education Research*, *1*(1), 40–95.

Crouch, C., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, *69*, 970–977.

Desrochers, M. N., & Shelnutt, J. M. (2012). Effect of answer format and review method on college students learning. *Computers and Education*, *59*, 946–951.

Elicker, J., & McConnell, N. (2011). Interactive learning in the classroom: Is student response method related to performance? *Teaching of Psychology*, *38*(3), 147–150.

Ertmer, P. A., Ottenbreit-Leftwich, A. T., Sadik, O., Sendurur, E., & Sendurur, P. (2012). Teachers beliefs and technology integration practices: A critical relationship. *Computers and Education*, *59*(2), 423–435.

Fortner-Wood, C., Armistead, L., Marchand, A., & Morris, F. B. (2013). The effects of student response systems on student learning and attitudes in undergraduate psychology courses. *Teaching of Psychology*, *40*(1), 26–30.

Gehlen-Baum, V., Weinberger, A., Pohl, A., & Bry, F. (2014). Technology use in lectures to enhance Students' attention. In C. Rensing, S. de Freitas, T. Ley, & P. J. Muñoz-Merino (Eds.), *Open learning and teaching in educational communities. EC-TEL 2014. Lecture notes in computer science, vol 8719*. Springer.

Hake, R. (1998). Interactive engagement versus traditional methods: A six-thousand student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64–74.

Han, J. H. (2014). Closing the missing links and opening the relationships among the factors: A literature review on the use of clicker technology using the 3P model. *Educational Technology and Society*, *17*(4), 150–168.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.

Hoekstra, A., & Mollborn, S. (2012). How clicker use facilitates existing pedagogical practices in higher education: Data from interdisciplinary research on student response systems. *Learning, Media and Technology*, *37*(3), 303–320.

Hunsu, N. J., Adesope, O., & Bayly, D. J. (2016). A meta-analysis of the effects of audience response systems (clicker-based technologies) on cognition and affect. *Computers & Education*, *94*, 102–119.

Kay, R. H., & LeSage, A. (2009). Examining the benefits and challenges of using audience response systems: A review of the literature. *Computer & Education*, *53*(3), 819–827.

Kennedy, G. E., & Cutts, Q. I. (2005). The association between students' use of electronic voting systems and their learning outcomes. *Journal of Computer Assisted Learning*, *21*(4), 260–268.

Kleitman, S., & Costa, D. S. J. (2014). The role of a novel formative assessment tool (stats-mIQ) and individual differences in real-life academic performance. *Learning and Individual Differences*, *29*, 150–161.

Koenig, K. (2010). Building acceptance for pedagogical reform through wide-scale implementation of clickers. *Journal of College Science Teaching*, *39*(3), 46–50.

Koretsky, M. D., Brooks, B. J., & Higgins, A. Z. (2016). Written justifications to multiple-choice concept questions during active learning in class. *International Journal of Science Education*, *38*(11), 1747–1765.

Lachner, A., Jacob, L., & Hoogerheide, V. (2021). Learning by writing explanations: Is explaining to a fictitious student more effective than self-explaining? *Learning and Instruction*, *74*, 101438.

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, *18*, 30–43.

Marshall, L., Valdosta, G., & Varnon, A. (2012). An empirical investigation of clicker technology in financial accounting principles. *Learning in Higher Education*, *8*(1), 7–17.

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., Bulger, M., Campbell, J., Knight, A., & Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, *34*(1), 51–57.

Mazur, E. (1997). *Peer instruction: A user's manual series in educational innovation*. Prentice Hall.

Mazur, E. (2009). Farewell, lecture? *Science*, *323*, 50–51.

McDonough, K., & Foote, J. (2015). The impact of individual and shared clicker use on students' collaborative learning. *Computers & Education*, *86*, 236–249.

Menary, R. (2007). Writing as thinking. *Language Sciences*, *29*, 621–632.

Michinov, N., Anquetil, E., & Michinov, E. (2020). Guiding the use of collective feedback displayed on heatmaps to reduce group conformity and improve learning in peer instruction. *Journal of Computer Assisted Learning*, *36*(6), 1026–1037.

Michinov, N., Morice, J., & Ferrières, V. (2015). A step further in peer instruction: Using the stepladder technique to improve learning. *Computers & Education*, *91*, 1–13.

Mollborn, S., & Hoekstra, A. (2010). "A meeting of minds": Using clickers for critical thinking and discussion in large sociology classes. *Teaching Sociology*, *38*, 18–27.

Mun, W. K., Hew, K. F., & Cheung, W. S. (2009). The impact of the use of response pad system on the learning of secondary school physics concepts: A Singapore quasi-experiment study. *British Journal of Educational Technology*, *40*(5), 848–860.

Nielsen, K. L., Hansen-Nygård, G., & Stav, J. B. (2012). Investigating peer instruction: How the initial voting session affects students' experiences of group discussion. *ISRN Education*, *2012*, 290157.

Nückles, M., Hubner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction*, *19*(3), 259–271.

Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J., & Renkl, A. (2020). The self-regulation-view in writing-to-learn: Using journal writing to optimize cognitive load in self-regulated learning. *Educational Psychology Review*, *32*(4), 1089–1126.

Papadopoulos, P. M., Demetriadis, S. N., Stamelos, I. G., & Tsoukalas, I. A. (2009). Prompting students' context-generating cognitive activity in ill-structured domains: does the prompting mode affect learning?. *Educational Technology Research and Development*, *57*(2), 193-210. http://dx.doi.org/10.1007/s11423-008-9105-6

Papadopoulos, P. M., Demetriadis, S. N., Stamelos, I. G., Tsoukalas, I. A. & Authors. (2011). The value of writing-to-learn when using question prompts to support web-based learning in ill-structured domains. *Educational Technology Research and Development*, *59*(1), 71-90. http://dx.doi.org/10.1007/s11423-010-9167-0

Papadopoulos, P. M., Natsis, A., Obwegeser, N., Weinberger, A. (2019). Enriching feedback in audience response systems: Analysis and implications of objective and subjective metrics on students' performance and attitudes. *Journal of Computer Assisted Learning*, *35*(2), 305-316. http://dx.doi.org/10.1111/jcal.12332

Papadopoulos, P. M., Lagkas, T. D., & Demetriadis, S. N. (2017). Technology-Enhanced Peer Review: Benefits and Implications of Providing Multiple Reviews. *Educational Technology and Society*, *20*(3), 69–81.

Perez, K. E., Strauss, E. A., Downey, N., Galbraith, A., Jeanne, R., Cooper, S., & Madison, W. (2010). Does displaying the class results affect student discussion during peer instruction? *CBE Life Science Education*, *9*, 133–140.

Poirier, C. R., & Feldman, R. S. (2007). Promoting active learning using individual response technology in large introductory psychology classes. *Teaching of Psychology*, *34*(3), 194–196.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, *93*(3), 223–231.

Schnaubert, L., & Bodemer, D. (2015). Subjective validity ratings to support shared knowledge construction in CSCL. In O. Lindwall, P. Häkkinen, T. Koschmann, P. Tchounikine, & S. Ludvigsen (Eds.), *Exploring the material conditions of learning: The computer supported collaborative learning (CSCL) conference 2015 (Vol. 2)* (pp. 933–934). International Society of the Learning Sciences.

Schnaubert, L., & Bodemer, D. (2017). Prompting and visualising monitoring outcomes: Guiding self-regulatory processes with confidence judgments. *Learning and Instruction*, *49*, 251–262.

Schnaubert, L., & Bodemer, D. (2019). Providing different types of group awareness information to guide collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, *14*, 7–51.

Schraw, G. (2009). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *The educational psychology series. Handbook of metacognition in education* (pp. 415–429). Routledge.

Shapiro, A. M. (2009). An empirical study of personal response technology for improving attendance and learning in a large class. *Journal of the Scholarship of Teaching and Learning*, *9*(1), 13–26.

Shapiro, A. M., & Gordon, L. (2013). Classroom clickers offer more than repetition: Converging evidence for the testing effect and confirmatory feedback in clicker-assisted learning. *Journal of Teaching and Learning with Technology*, *2*(1), 15–30.

Shapiro, A. M., & Gordon, L. T. (2012). A controlled study of clicker-assisted memory enhancement in college classrooms. *Applied Cognitive Psychology*, *26*, 635–643.

Shapiro, A. M., Sims-Knight, J., O'Rielly, G. V., Capaldo, P., Pedlow, T., Gordon, L., & Monteiro, K. (2017). Clickers can promote fact retention but impede conceptual understanding: The effect of the interaction between clicker use and pedagogy on learning. *Computers & Education*, *111*, 44–59.

Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 553–558.

Stankov, L. (2013). Noncognitive predictors of intelligence and academic achievement: An important role of confidence. *Personality and Individual Differences*, *55*(7), 727–732.

Stankov, L., Morony, S., & Lee, Y. P. (2013). Confidence: The best non-cognitive predictor of academic achievement? *Educational Psychology*, *34*, 9–28.

Stowell, J., Oldham, T., & Bennett, D. (2010). Using student response systems ("clickers") to combat conformity and shyness. *Teaching of Psychology*, *37*(2), 135–140.

Strasser, N. (2010). Who wants to pass math? Using clickers in calculus. *Journal of College Teaching & Learning*, *7*(3), 49–52.

Sutherlin, A. L., Sutherlin, G. R., & Akpanudo, U. (2013). The effect of clickers in university science courses. *Journal of Science Education and Technology*, *22*(5), 651–666.

Veenman, M. V. J. (2017). Learning to self-monitor and self-regulate. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (2nd ed., pp. 233–257). Routledge.

Vickrey, T., Rosploch, K., Rahmanian, R., Pilarz, M., & Stains, M. (2015). Research-based implementation of peer instruction: A literature review. *CBE Life Sciences Education*, *14*(1), es3.

Wang, Y., Lin, L., & Chen, O. (2021). The benefits of teaching on comprehension, motivation, and perceived difficulty: Empirical evidence of teaching expectancy and the interactivity of teaching. *The British Journal of Educational Psychology*, 12416. https://doi.org/10.1111/bjep.12416

Wu, Y.-C. J., Wu, T., & Li, Y. (2019). Impact of using classroom response systems on students' entrepreneurship learning experience. *Computers in Human Behavior*, *92*, 643–645.

How to cite this article: Papadopoulos, P. M., Obwegeser, N., & Weinberger, A. (2021). Let me explain! The effects of writing and reading short justifications on students' performance, confidence and opinions in audience response systems. *Journal of Computer Assisted Learning*, 1–11. https://doi.org/10.1111/jcal.12608